

AD-A259 081



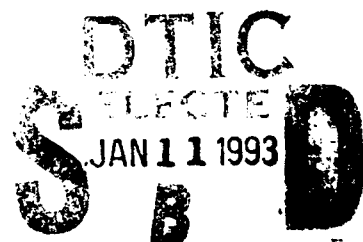
AFIT/GE/ENG/92D-38

Signal Approximation with a Wavelet Neural Network

THESIS

Charles M. Westphal

AFIT/GE/ENG/92D-38



012233
93-00097



Approved for public release; distribution unlimited

AFIT/GE/ENG/92D-38

Signal Approximation with a Wavelet Neural Network

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Electrical Engineering

DTIC QUALITY ASSURANCE

Charles M. Westphal, B.S.

December, 1992

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Special and/or
A-1	Special

Approved for public release: distribution unlimited

Table of Contents

	Page
Table of Contents	ii
List of Figures	v
List of Tables	vi
Abstract	vii
 I. PROBLEM STATEMENT	 1
1.1 Introduction	1
1.2 Background	1
1.2.1 Artificial Neural Networks	1
1.2.2 Wavelet Transform	2
1.3 Summary of Current Knowledge	3
1.4 Problem	3
1.5 Scope	4
1.6 Approach/Methodology	4
1.7 Conclusion	4
 II. LITERATURE REVIEW	 5
2.1 Introduction	5
2.2 Artificial Neural Networks	5
2.2.1 Function Approximation and Hidden-Layer Nodes . . .	5
2.2.2 Hyperplane and Kernel Classifier Neural Networks . . .	5
2.3 Wavelets	6
2.3.1 Time-Frequency Resolution	6
2.3.2 Filter Banks	6

	Page
2.4 Wavelet Neural Networks	6
2.5 Conclusion	7
III. THEORY AND PROCEDURE	9
3.1 Introduction	9
3.2 Artificial Neural Networks (ANNs)	9
3.2.1 Linear ANNs	9
3.2.2 Non-Linear ANNs	11
3.2.3 Feedforward ANNs	12
3.3 Wavelet Analysis	12
3.3.1 Fourier Analysis	12
3.3.2 Short-Time Fourier Transform (STFT)	14
3.3.3 Wavelet Transform	15
3.3.4 Frames	17
3.4 Wave-Nets	21
3.5 Summary	23
IV. RESULTS	24
4.1 Sigmoidal Wavelet	24
4.2 Wavelet Neural Network	27
4.2.1 Network Architecture	27
4.2.2 Software Implementation	30
4.2.3 Sampling of Training and Testing Data	30
4.2.4 Approximation Results	30
4.3 Hardware Implementation	33
V. CONCLUSIONS AND RECOMMENDATIONS	39
5.1 Conclusions	39
5.2 Recommendations	39

	Page
Appendix A. Fourier Transforms of Sigmoidal Wavelets	40
Bibliography	46
Vita	48

List of Figures

Figure		Page
1.	Single Layer Perceptron	10
2.	Sigmoid and Hyperbolic Tangent Functions	11
3.	Time-frequency plane of the STFT	16
4.	Single-Input, Single Output ANN Architecture	22
5.	Combination of Two Sigmoidal Functions	24
6.	Candidate Sigmoidal Mother Wavelet	25
7.	Fourier Transform of Candidate Sigmoidal Mother Wavelet, $\Psi(\omega)$	26
8.	Time-Frequency Region for Approximation	28
9.	Approximation of Two Sinusoids	31
10.	Approximation of Random Frequencies between 5 and 200 Hz	32
11.	Sampled Signal of a Pulsed Carrier Frequency	33
12.	Envelope of Typical Signal After Hilbert Transform	34
13.	Approximation of Signal with Varying Numbers of Dilations	35
14.	Sigmoidal Wavelet Functions	36
15.	Magnitude of FT of Symmetric Sigmoidal Wavelet	37
16.	Hardware Implementation for Sigmoidal Wavelets	38
17.	Integration Contour	42

List of Tables

Table	Page
1. Frequency Domain Parameters for Example	29
2. Time Domain Parameters for Example	29

Abstract

This study investigated the use of Wavelet Neural Networks (WNN) for signal approximation. The particular wavelet function used in this analysis consisted of a summation of sigmoidal functions (a *sigmoidal wavelet*). The sigmoidal wavelet has the advantage of being easily implemented in hardware via specialized electronic devices like the Intel Electronically Trainable Analog Neural Network (ETANN) chip. The WNN representation allows the determination of the number of hidden-layer nodes required to achieve a desired level of approximation accuracy. Results show that a bandlimited signal can be accurately approximated with a WNN trained with irregularly sampled data.

Signal Approximation with a Wavelet Neural Network

I. PROBLEM STATEMENT

1.1 Introduction

Maintaining an accurate assessment of the electromagnetic environment expected in modern conflicts is a challenging task. Part of the problem lies in the large number of signals occupying the same frequency spectrum making the job of detecting and tracking specific emitters difficult. Real-time processing of the resulting data could exceed the processing capabilities of current monitoring systems. Another concern based on today's technology is the possibility of a system time-varying its emitted signal. Any time variation in the signal significantly complicates the task of tracking such an emitter using traditional detection and identification techniques. A promising development in solving the real-time data processing problem is the artificial neural network (ANN). Because of the parallel architecture of an ANN, its theoretical capability to process data is much higher than conventional signal processing techniques. An ANN does not, by itself, solve the problem of tracking time varying signals, however, a popular area of recent research in the analysis of time-varying signals is the Wavelet Transform (WT). The combination of ANNs and the WT presents a possible solution to this overall problem. This thesis examines the implementation of a Wavelet-based ANN for performing signal approximations as a first step toward ultimately performing signal detection, identification and classification with a Wavelet Neural Network (WNN).

1.2 Background

This section provides a short background in the areas of artificial neural networks and the wavelet transform.

1.2.1 Artificial Neural Networks ANNs are based on simple models of the biological neuron and have been widely studied in many areas of science and engineering. Like

the highly intertwined neuron cells, a general ANN architecture can be highly interconnected and consist of many layers of interconnected nodes. Node outputs can be calculated independently in the same layer, and thus the claim of a parallel architecture. The other important characteristic of neural networks modeled on biological neurons is the ability to generalize. The generalization property of ANNs refers to "the ability of the network to use its training data to deal with the real world events not in the training data." (13)

Neural networks are usually trained using one of two general methods: supervised training where training data is labeled as to which class it belongs, and unsupervised training where the neural network arranges its nodes based on the unlabeled training data. The primary interest for this thesis is the supervised training algorithms. A wide variety of applications have been analyzed using ANNs including radar emitter identification, speech recognition, and stock market prediction, to name just a few (13). Several problems that have proven extremely difficult using conventional analytical techniques have been solvable using ANNs. Research efforts at the Air Force Institute of Technology (AFIT) on identifying and classifying electronic signals with neural networks are reported in (4) and (17).

1.2.2 Wavelet Transform Several important signal analysis techniques, including multiresolution analysis, can be represented in terms of the WT. The basic idea of the WT is similar to the Short Time Fourier Transform (STFT) or Windowed Fourier Transform (WFT). The WT and STFT both "slice" a signal into smaller sections and operate on each section separately. The major differences between the WT and STFT are that the kernel function for the WT is much less restricted (not limited to a specific function like the complex exponential of the Fourier Transform) and the window function of the WT varies in size. The varying window of the WT is assumed to be an advantage because real-world time-varying signals can be composed of relatively long-lived low frequency components and relatively short-lived high frequency components which may be difficult to analyze with the fixed window of the STFT. Because there are relatively few restrictions on the kernel function of the WT, many wavelet kernel functions are possible and it is generally suggested to choose a kernel function which "looks" like the functions being analyzed.

Some of the promising areas of research involving wavelets have been in the areas of data compression and image processing and the topic has generated considerable interest in many fields of study.

1.3 Summary of Current Knowledge

The fundamental theory of the WNN on which this research is based has been reported by Pati and Krishnaprasad (11). Their work involves a neural network that performs function approximation and demonstrates a neural network based on wavelet transformations has several desirable properties. These properties will be discussed further in Chapter 2, but the basic ideas are listed here: 1) the number of nodes in the single hidden layer of the network can be accurately computed (this is generally selected heuristically in most neural networks) and 2) two-thirds of the weights of the neural network are precomputed and set to constants so that training only involves the layer of weights from the hidden layer to the output layer. The weights from the input to the hidden layer of the WNN are predetermined based on the chosen wavelet transform as well as the bandwidth and time-bandwidth product of the signal being analyzed. The weights from the hidden layer to the output layer are learned and represent the projection coefficients of the signal onto the wavelet function. Another aspect of (11) to be continued in this thesis is the particular neural network architecture used that is easily implemented in hardware. Choosing a wavelet function composed of a summation of sigmoidal functions allows the use of new computer chips that function as artificial neural networks with sigmoidal activation functions (3).

1.4 Problem

The basic problem to be analyzed for this thesis is to approximate real-world time-varying electronic signals with a WNN. The goal is perform initial work that could lead to ultimately developing a neural network architecture that sufficiently characterizes time-varying electronic signals for the tasks of detection, identification, and classification.

1.5 Scope

This research is limited to developing a WNN in software and to evaluating its performance in approximating real-world data. The particular wavelet transform chosen to be implemented in this study, composed of summing sigmoidal functions, has the advantage of being easily implemented in hardware. The data set used for testing the usefulness of the wavelet neural network is real-world data, obtained from the Avionics Directorate of Wright Laboratories, Wright-Patterson AFB, OH.

1.6 Approach/Methodology

The first step in this thesis will be to review the theory of WNNs. The next step will be the implementation of a WNN in software that performs function approximation. This will be followed by investigating the performance of the WNN in approximating real-world signals.

1.7 Conclusion

The order of presentation of this thesis will be first to review the literature in the areas of artificial neural networks, wavelets, and WNNs in Chapter II. Next are explanations of neural networks, wavelets, and WNNs as they pertain to the problem statement to be discussed in Chapter III. Chapter IV contains the results achieved in this effort. Chapter V contains the conclusions and recommendations for further research in this area.

II. LITERATURE REVIEW

2.1 Introduction

This chapter reviews recent literature in the area of combining the topics of artificial neural networks and wavelets for the purpose of function approximation and prediction. The combined subject is referred to as Wavelet Neural Networks (WNNs) in this thesis and as *Wave-Nets* by one of the authors in this field (1).

2.2 Artificial Neural Networks

2.2.1 Function Approximation and Hidden-Layer Nodes A single hidden layer network with sigmoidal activation functions is all that is needed, given sufficient nodes, to perform any arbitrary transformation as shown by Cybenko (5); however, multiple layers are generally used because of faster training (13). The ANN configuration necessary to achieve a "close" approximation of a function may require a very large number of nodes. In general, no technique exists for determining the number of hidden-layer nodes required for a given level of performance. The experimenter usually starts with a small number of nodes in the hidden-layer and increases the number until the desired level of performance is achieved.

2.2.2 Hyperplane and Kernel Classifier Neural Networks As summarized by Zahirniak (17), both Hyperplane and Kernel Classifier neural networks can approximate any multivariate function based on the Cybenko theory. Hyperplane Classifier neural networks generally use superposition of sigmoidal functions to achieve an approximation of a function. The Hyperplane Classifier is also referred to as a global classifier because the decision region is not restricted to the training data and may over-generalize. Kernel Classifier neural networks, such as the radial basis function classifier, use overlapping kernel functions that create local decision regions. The Kernel Classifier trains faster than the Hyperplane Classifier but may not generalize well for data unlike that used in the training set.

2.3 Wavelets

2.3.1 Time-Frequency Resolution A principal motivation for using the WT is its ability to examine transient signal phenomena not possible with the Short-Time Fourier transform (STFT), the classical method for studying non-stationary signals (6). The varying time-frequency resolution of the WT contrasts with the time-frequency resolution of the STFT which has a fixed resolution for all frequencies. The justification for using the wavelet transform over the STFT is because signals of practical interest may be composed of short-duration, high-frequency components and long-duration, low-frequency components (12).

2.3.2 Filter Banks Filter banks and multiresolution signal analysis were theoretically linked to wavelet analysis by Mallat (9). In terms of signal analysis, it is often convenient to think of wavelet analysis in terms of a "constant-Q" bandpass filter bank (12). Constant-Q, or constant relative bandwidth, filters can be found in biological processes like the auditory system (16).

2.4 Wavelet Neural Networks

As discussed above, Hyperplane Classifier ANNs are typically constructed using sigmoidal functions (or hyperbolic tangent functions) as non-linear activation functions. By representing a feedforward ANN as groups of summations of sigmoidal activated nodes to form wavelet transformations, Pati and Krishnaprasad have developed a function approximator that uses sigmoidal activation functions but represents a Kernel Classifier (11). Unlike most ANN architectures, it was shown in this paper a method for determining the number of nodes required in the hidden layer of a two-layer ANN by using the bandwidth of the signal and the time-bandwidth product to achieve a desired level of approximation. Because the wavelet basis functions used in this article were constructed by summing sigmoid functions this architecture has the advantage of being easily implemented in hardware.

Work by Bakshi and Stephanopoulos showed results similar to (11) but used an ANN with orthonormal wavelet activation functions (1). The authors use the term "Wave-Net" to describe the resulting ANN which is a feedforward neural network with a single

hidden layer. A coarse approximation is first achieved using nodes with the wavelet scaling functions as activation functions and refined approximations are achieved by adding nodes with the appropriate wavelet activation functions. Using orthogonal wavelets as basis functions, rather than other standard orthogonal basis functions, has the advantage that wavelets are localized in time and frequency. This localization is important in terms of neural networks because global (non-localized) activation functions have been found to train slowly and do not guarantee convergence. The other advantage to using wavelets is the multiresolution property of the function representation, meaning variable resolutions of the input data are permissible while still explicitly representing the function. However, the use of orthonormal wavelets would not be as easily implemented in hardware as the wavelets used in (11).

From Mallat's work, it is known that the approximation of a function using wavelets can be represented in terms of the projection and the detail at a particular resolution (9). In the hidden-layer of the Wave-Net, the projections are obtained using what the authors term "scaling function nodes" and the details are obtained from the "wavelet nodes" (1). This representation is somewhat different than (11) where non-orthogonal wavelets are used, no detail coefficients are computed, and only projection coefficients are learned. The method presented in (1) is to use the coarsest approximation possible for the scaling function nodes and train the network until the training data are "overfitted" (this means to train until the error starts to increase). The approximation error is further reduced by adding the wavelet nodes until the desired approximation error is achieved. Wavelet nodes with small weights can be eliminated to reduce the dimension of the resulting network. The methods of (11) and (1) both allow the input data be irregularly sampled, i.e., the training data can be sampled at a different rate than the test data and the function. Irregularly sampled data is not as well accommodated in other types of wavelet analysis such as the Mallat algorithm.

2.5 Conclusion

This chapter reviewed recent work in the areas of ANNs and wavelet analysis as related to the topic of WNNs in performing function approximation and prediction. The

wavelet theory applied to ANNs provides a method for calculating the number of hidden layer nodes required to achieve a desired level of approximation accuracy in the function approximation. Other benefits of the WNNs discussed are the use of sigmoidal activation functions which can be easily implemented in hardware and approximating a function with irregularly sampled training data. The work of Pati and Krishnaprasad will form the foundation of the the thesis. In particular, the use of a wavelet composed of summing sigmoidal functions will be maintained for the purpose hardware implementation.

III. THEORY AND PROCEDURE

3.1 Introduction

This chapter introduces the topics of artificial neural networks (ANNs) and wavelets as they apply to the approximation and classification of time-varying signals. As described in Chapter II, it has been shown that wavelet analysis can be used to describe the process by which the ANN approximates a function. This chapter begins with a short description of linear and non-linear ANNs and the typical feed-forward ANN. The discussion then shifts to the Wavelet Transform by first introducing the Fourier Transform (FT) and the Short-Time Fourier Transform (STFT). Of importance in this area is the idea of using *basis functions*, or “building blocks” to represent signals which allows relevant information about the signal to be more easily examined and may dramatically simplify certain mathematical operations. Time-frequency analysis is the other major idea covered in this section on the WT. Overcoming the lack of time information in the FT and the fixed time-frequency resolution of the STFT was the main motivation for the development of the WT.

3.2 Artificial Neural Networks (ANNs)

3.2.1 Linear ANNs The single-layer ANN is referred to as the perceptron from Widrow’s work in the early 1960’s (13) and is represented by the diagram in Figure 1. Each element of the input vector \bar{x} is multiplied by a weight. When the activation function of the ANN is linear, then the node output is simply the scaled sum of the weighted inputs to the node. The resulting network can be represented in terms of a matrix equation. Considering the activation function $f(x) = x$, the input x , and the output y , then the equation for the linear network can be written as

$$y = Ax \tag{1}$$

where A is the matrix of weights $\{w_{11}, w_{21}, \dots, w_{ij}\}$ for i inputs and j outputs. Using matrix algebra, it is possible to find the psuedo-inverse of A denoted as \acute{A} . \acute{A} has the following

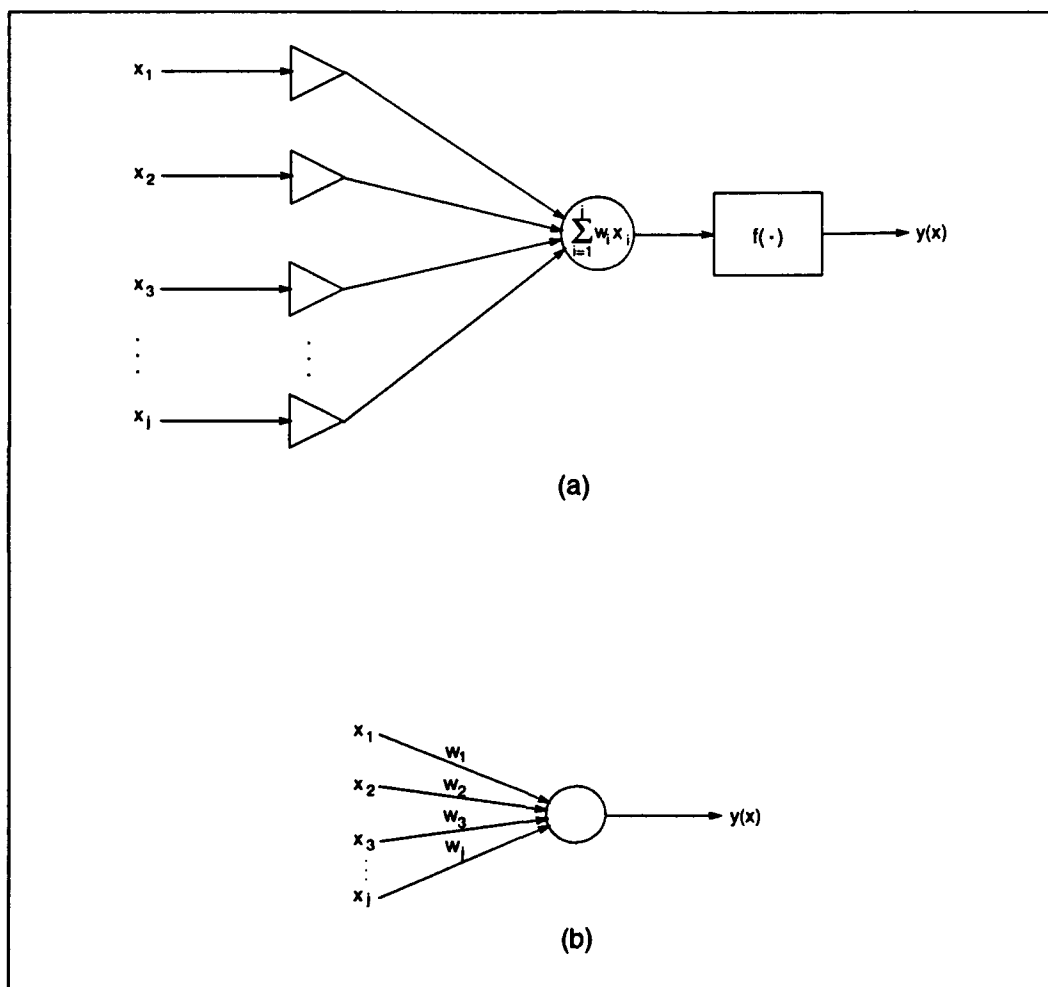


Figure 1. Single Layer Perceptron. (a) Actual configuration (b) Short-hand Notation Version

properties (15):

$$\dot{A} = A^{-1} \quad \text{if } A \text{ is square and non-singular} \quad (2)$$

$$\dot{A}A = I \quad \text{where } I \text{ is the identity matrix} \quad (3)$$

$$\dot{A} = (A^T A)^{-1} A^T \quad \text{if } A \text{ has full column rank} \quad (4)$$

3.2.2 Non-Linear ANNs Non-linear activation functions are commonly used which model biological neurons in that the output has a maximum value and is limited to some range. A popular choice for the activation function is the so-called sigmoid function which is defined as:

$$s(t) = \frac{1}{1 + e^{-t}} \quad (5)$$

Another typical non-linear functions used as an activation function is the hyperbolic tangent function, $\tanh(t)$, which can be viewed as a sigmoid function shifted and scaled along the y-axis. A plot of the sigmoid and hyperbolic tangent functions is provided in Figure 2. Besides better modeling biology, the non-linear ANN provides solutions to a more general set of problems than the linear ANN.

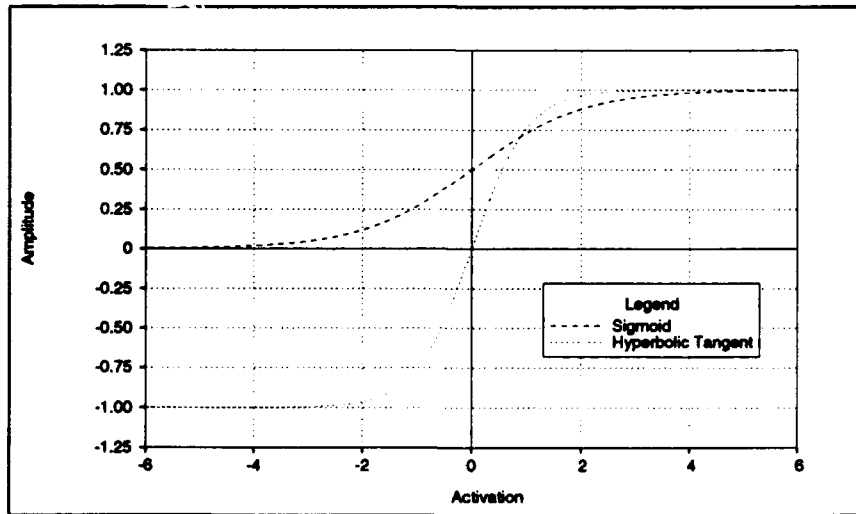


Figure 2. Sigmoid and Hyperbolic Tangent Functions

3.2.3 Feedforward ANNs A feedforward neural network is an ANN in which all the connections are made in the direction from the input toward the output. The common method for training the feedforward neural network is referred to as the backward propagation algorithm or just backprop. Backprop training is a method for adjusting the weights of the ANN beginning with the weights of the output layer and working backward toward the input layer. Modification of the weights is accomplished by calculating the output of the ANN for a given input and comparing the actual output with the desired output. The weights of the input to the output layer are then modified to minimize the cost function that represents the difference between the actual and desired outputs. The procedure is repeated for each layer of nodes in the network until the weights of the input layer are adjusted. The backprop algorithm is described in (13). It has been shown that the ANN is actually learning the Bayesian classification which removes some of the “black box” mystique of the ANN solution and how it is achieved (14).

3.3 Wavelet Analysis

The goal of this section is to introduce the topic of wavelet analysis and a few of the important differences of the wavelet transform with the standard time-frequency analysis techniques. The discussion begins with the basic tool of signal analysis, the Fourier Transform (FT).

3.3.1 Fourier Analysis The spectrum, or frequency content, of a signal is found by taking the FT. The FT pair is defined as the following:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt \quad (6)$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega)e^{j\omega t} d\omega \quad (7)$$

Most real-world signals meet the conditions required for using the FT (as long as the signal is absolutely squared-integrable on the infinite interval $(-\infty, +\infty)$ (7). Mathematically, the FT provides a description of the signal in the frequency plane by representing

the signal in terms of an infinite number of weighted complex exponential *basis functions* (sometimes referred to as *analyzing functions*).

Closely related to the FT is the Fourier Series (FS) for periodic signals which is defined as

$$f(n) = \sum_{n \in \mathbb{Z}} c_n e^{-j\omega_0 n} \quad (8)$$

The basis functions for the FS are also the complex exponential functions. Any well-behaved function can be represented in terms of an infinite sum of complex exponentials, which form an orthonormal basis set.

3.3.1.1 Properties of the FT The magnitude and phase of the FT can be represented using the notation

$$F(\omega) = A(\omega)e^{-j\Theta(\omega)} \quad (9)$$

The magnitude of the FT is an even function and the phase is an odd function.

$$A(\omega) = A(-\omega) \quad (10)$$

$$\Theta(\omega) = -\Theta(-\omega) \quad (11)$$

Also useful in the discussion to follow is Parseval's Identity which states the signal energy in the time domain equals the spectrum energy in the frequency domain.

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = 2\pi \int_{-\infty}^{+\infty} |F(\omega)|^2 d\omega \quad (12)$$

The complex exponentials used as the basis functions in the FT are infinite in duration which makes the Fourier transform unable to provide information about when a particular frequency component occurs in the signal. That is, the Fourier transform is a convenient analytical tool when the signal of interest is stationary (not time-varying). If the signal of interest is time-varying, a particular frequency component of the FT may

occur at any time during the period of analysis. The time of occurrence of particular frequency components can be extremely important in many types of signal analysis. As an example, a musical score provides the timing and length of notes as well as the particular notes to use. Without both the time and frequency information, a musical score would not be of much use to the musician.

3.3.2 Short-Time Fourier Transform (STFT) The first technique developed to provide time-frequency information is referred to as the Short-Time Fourier Transform (STFT) or Windowed Fourier Transform (WFT). Time localization is achieved by windowing or “slicing” the signal of interest into small intervals of the total time and taking the FT of each interval. Shifting the window changes the interval of time in which the FT is taken. Alternatively, the STFT can be viewed as taking the infinite duration complex exponentials of the FT (which provide the frequency localization) and multiplying by a finite duration window function (which provides the time localization). Thus, the STFT provides both time and frequency localization. The STFT is the standard technique for time-frequency analysis and is defined as the following:

$$STFT(\omega) = \int_{-\infty}^{+\infty} f(t)g^*(t - \tau)e^{-j\omega t} dt \quad (13)$$

where $g(t)$ is the window function. Various window functions are used depending on the application with some of the more common windows being the Hamming and Hanning (10). When a Gaussian windowing function is chosen, the STFT is referred to as the Gabor transform which has received considerable interest in research. Because the Fourier Transform of a Gaussian function is another Gaussian function, the Gabor transform performs Gaussian windowing in both time and frequency domains and is well-localized in both time and frequency. Other windows well localized in time and frequency are possible. If both the time and frequency domain windows are concentrated around zero, then the STFT can be “loosely” interpreted as the “content” of the signal near a time t and near a frequency ω (6).

From a mathematical perspective, the STFT can be represented in terms of basis functions where the finite duration of each time interval requires a summation of translated

windows to “cover” the entire time line. Often, the overlap of the windows is chosen to minimize the resulting distortion to the original signal. An alternative view to considering the STFT as representing a signal in terms of weighted basis functions is to consider it a modulated filter bank. The filter bank approach, developed in digital signal processing, represents the STFT as group of filters whose frequency range is determined by the window function $g(t)$. Which of the two views to choose when using the STFT depends on the type of analysis being performed and the properties of the STFT that are being exploited. Important properties to be considered are discussed in the following section.

3.3.2.1 Properties of STFT The window size of the STFT is fixed; thus, there is a lower bound on the time-frequency resolution that can be achieved. The uncertainty principle states the time-bandwidth product, $\Delta_t \Delta_\omega$ must be

$$\Delta_t \Delta_\omega \geq \frac{1}{4\pi} \quad (14)$$

with the Gabor Transform (Gaussian window) providing the minimum time-frequency resolution. When analyzing signals where various frequency components are of interest, a frequency resolution of Δ_ω may be sufficient for one frequency component in the signal, but not for others. The STFT can be considered to have a fixed window length in time that looks at all frequencies, or as a fixed window length in frequency that looks at all time as shown in Figure 3. What is desired is a set of basis functions where the time-frequency resolution varies; a variable time-frequency resolution is more representative of real-world signals in which low-frequency components are relatively long lived, while high-frequency components are relatively short lived. The solution to the variable time-frequency resolution is the Wavelet Transform (WT). Although both the STFT and the WT provide a description of f (the signal of interest) in the time-frequency plane, the WT provides a similar time-frequency description with a few important differences (6) which are discussed in the next section.

3.3.3 Wavelet Transform Whereas the STFT has a fixed window length that is translated to the proper time location and results in a fixed frequency resolution, the WT

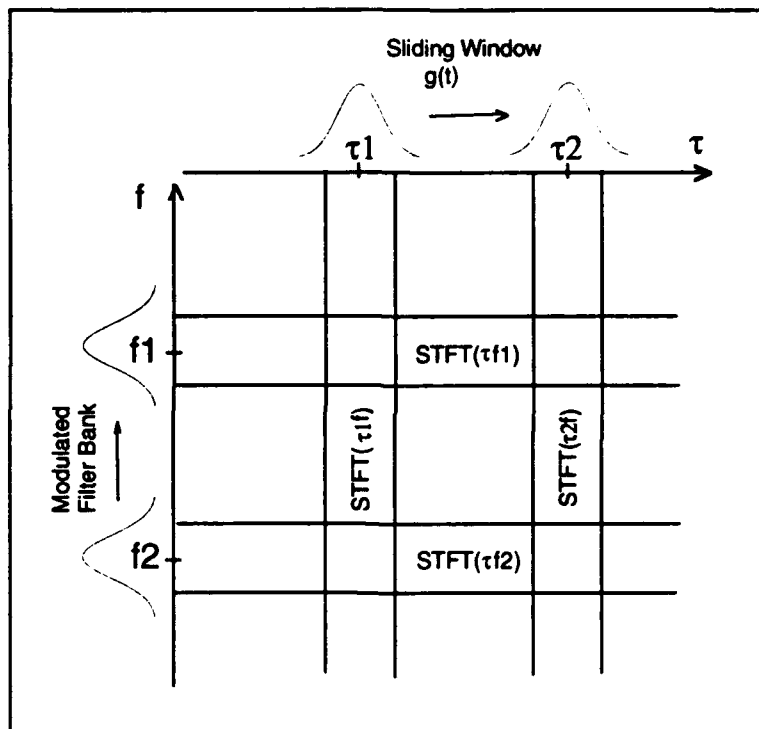


Figure 3. Time-frequency plane of the STFT (12)

has a variable width window which can be narrow in time for high frequencies and wide in time for low frequencies. As previously mentioned, the variable window width of the WT better matches time-varying signals. The first requirement in using the WT is to find a valid *mother* or *basic* wavelet which has windowing properties in both the time and frequency domains. The variable time-frequency resolution of the wavelet transform is achieved by dilating and translating the basic wavelet function, $\psi(t)$.

$$\psi_{ab}(t) = \psi\left(\frac{t-b}{a}\right) \quad (15)$$

where a is the dilation parameter and b is the translation parameter. The WT uses logarithmic shifts of the mother wavelet.

Typically, the dilation parameter, a , is chosen to be 2 resulting in octave dilations and the translation parameter, b is taken to be integers; the Continuous Wavelet Transform is then defined as

$$\mathcal{W}_{mn}(a, b) = 2^{-\frac{n}{2}} \int_{-\infty}^{+\infty} f(t) 2^{-n} \psi(t - mb) dt \quad (16)$$

The *admissability* condition for a function, $\psi(t)$ to be called a *mother wavelet* is that the Fourier transform of $\psi(t)$, $\Psi(\omega)$, satisfy

$$\int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega = C_v < \infty \quad (17)$$

The admissability condition implies that the wavelet function, $\psi(t)$ meet the condition of zero average energy

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (18)$$

The characteristic shape of the time-domain plot of wavelet functions led to the naming of this class of functions as "wavelet" by a group of French researchers (12).

3.3.4 Frames A candidate mother wavelet does not necessarily form a frame. The necessity to generate a frame occurs due to the requirement to "cover" the entire line in the

time and frequency domains. If each translation of the dilated mother wavelet is shifted too far, then “holes” could develop in the coverage. Daubechies presented the necessary conditions to ensure a set of wavelet functions formed a frame (6). The first step in constructing a frame from the candidate mother wavelet is to describe the time-frequency localization of the function. The first parameter of interest is the center of localization in both the time and frequency domains. In the time domain, the center of the localization can be defined as

$$t_c = \frac{1}{\|f\|^2} \int_{-\infty}^{\infty} t |f(t)|^2 dt \quad (19)$$

Similarly, the center of localization in the frequency domain is defined as

$$\omega_c = \frac{2}{\|F\|^2} \int_0^{\infty} \omega |F(\omega)|^2 d\omega \quad (20)$$

where only the positive center frequency is needed because the magnitude of the Fourier Transform is an even function. Using Parseval's theorem,

$$\|f\|^2 = \frac{1}{2\pi} \|F\|^2 \quad (21)$$

the positive center frequency can be rewritten as

$$\omega_c = \frac{1}{\pi \|f\|^2} \int_0^{\infty} \omega |F(\omega)|^2 d\omega \quad (22)$$

The support of a function f is denoted by $\text{supp}(f)$ which represents the set for which the absolute value of $f(t)$ is greater than zero, or mathematically, the closure set $\{t : |f(t)| > 0\}$. If the function is always nonzero (such as a Gaussian function), then the area of localization for both time and frequency domain can be defined as an epsilon-support where the function gets arbitrarily close to zero, $\{t : |f(t)| > \epsilon\}$. Because the support of some candidate mother wavelets is always non-zero over $(-\infty, +\infty)$, an alternate definition is used called the epsilon energy support. The epsilon-support is a method for defining the interval over which the energy in the signal is contained except for an arbitrarily small epsilon amount, where epsilon represents a small number. Because the candidate mother

wavelet is assumed an odd function in the time domain (can always be made to be an odd function since a mother wavelet must satisfy the zero-average energy condition), the center of concentration in time is zero such that $t_c - t_0 = t_c + t_1$, or $t_0 = -t_1$. One way of defining the interval over which a function is concentrated in terms of its energy is to find the energy contained inside the interval $[t_0, t_1]$ symmetric about the center of concentration, t_c resulting in

$$\frac{1}{\|f\|^2} \int_{-\infty}^{t_0} |f(t)|^2 dt + \frac{1}{\|f\|^2} \int_{t_1}^{\infty} |f(t)|^2 dt < \epsilon \quad (23)$$

$$\frac{1}{\|f\|^2} \int_{t_0}^{t_1} |f(t)|^2 dt > 1 - \epsilon \quad (24)$$

What is sought is the smallest interval containing $(1 - \epsilon)$ of the total energy of the function. The epsilon support will be used to approximate the function to a desired level of accuracy. As $\epsilon \rightarrow 0$, the approximation of the function, $\tilde{f}(t)$, will approach $f(t)$ (in terms of energy) with the difference being arbitrarily small. The interval defined by a particular epsilon support gets larger as epsilon gets smaller (inverse relationship provided f has infinite extent). The epsilon energy support is defined as the smallest interval $[t_0, t_1]$ which satisfies

$$\epsilon_t\text{-supp}(\psi) = \frac{1}{\|\psi\|^2} \int_{t_0}^{t_1} |\psi(t)|^2 dt > 1 - \epsilon_t \quad (25)$$

In the frequency domain, a somewhat different definition for the epsilon energy support is required. The idea is to find the symmetric interval $[\omega_0, \omega_1]$ around the center frequency such that

$$\frac{1}{\|f\|^2} \int_{\omega_0}^{\omega_1} |F(\omega)|^2 d\omega > 1 - \epsilon_\omega \quad (26)$$

where ω_0 and ω_1 are equally spaced around ω_c and ω_0 is the maximum of $[0, \omega_0]$. Only the positive frequencies are considered which results in an epsilon energy support of the

smallest interval of $[\omega_0, \omega_1]$ which satisfies

$$\epsilon_\omega\text{-supp}(\psi) = \frac{1}{2 \|\psi\|^2} \int_{\omega_0}^{\omega_1} |\Psi(t)|^2 df > 1 - \epsilon_\omega \quad (27)$$

With the epsilon support defined, it is possible to continue the task of defining a frame. In general, the dilations and translations of the mother wavelet create a set of functions $[(\psi, a, b)]$ which is written as

$$\psi_{mn} = a^{\frac{n}{2}} \psi(a^n t - mb) \quad (28)$$

where m is an integer and n is any positive integer. As previously noted, dyadic dilations are generally used ($a = 2$) which results in the dilations being separated by octaves

$$\psi_{mn} = 2^{\frac{n}{2}} \psi(2^n t - mb) \quad (29)$$

What is not known at this point is whether a frame can be generated by the mother wavelet. The mathematics for generating frames require that

$$m(\psi, a) = \inf \left\{ \sum_n |\tilde{\psi}(a^n \omega)|^2 : |\omega| \in [1, a] \right\} > 0 \quad (30)$$

$$M(\psi, a) = \sup \left\{ \sum_n |\tilde{\psi}(a^n \omega)|^2 : |\omega| \in [1, a] \right\} < \infty \quad (31)$$

$$\lim_{b \rightarrow 0} 2 \sum_{k=1}^{\infty} \beta(k/b)^{\frac{1}{2}} \beta(-k/b)^{\frac{1}{2}} = 0 \quad (32)$$

where

$$\beta(s) = \sup \left\{ \sum_n |\hat{g}(a^n \omega)| |\hat{g}(a^n \omega - s)| : |\omega| \in [1, a] \right\} \quad (33)$$

where \inf is the infimum (largest lower bound) and \sup is the supremum operator (smallest upper bound) (6). Assuming a frame does exist, the next step is a matter of calculating

$$\sum_n |\hat{g}(a^n \omega)|^2 \text{ for } \omega \in [1, a] \quad (34)$$

and plotting

$$2 \sum_{k \neq 0} \beta(k/b)^{1/2} \beta(-k/b)^{1/2} \quad (35)$$

for various values of b as shown in (11). The translation parameter b is selected as the value which makes Equation 35 equal the result of Equation 34.

The dilated and translated functions defined in Equation (29) are *affine wavelets* where the term *affine* refers to the fact the transformation does not map zero in the time domain to zero in the wavelet transform domain. The wavelet transformation used here is termed *discrete* because the continuous function $\psi(t)$ is dilated at integer multiples n . A finite length, bandlimited signal to be analyzed will only exist over part of the time-frequency space defined by examining the frequency spectrum and time-bandwidth product of the signal. The bandwidth of the signal determines the required dilations of the mother wavelet needed to represent the signal. The time-bandwidth product provides the required translations.

3.4 Wave-Nets

The discrete wavelet representation as described by

$$f(t) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}^+} c_{mn} \psi_{mn}(t) \quad (36)$$

can be computed in a ANN which has a single input and single output (SISO). The basic architecture is shown in Figure 4. As discussed in (11), the integer dilations and translations can be predetermined based on the frequency bandwidth and the time-bandwidth product of the signal $f(t)$. A solution to finding these coefficients is to have a ANN "learn" them. Using a gradient descent backpropagation algorithm, an ANN can be trained to find the coefficients within some desired level of accuracy. The basic idea is that the SISO ANN accepts a time value as an input with the output an approximation of the original signal at the input time. During training, the coefficients are adjusted up or down so that the approximated value approaches the actual value. The approximation to the signal can

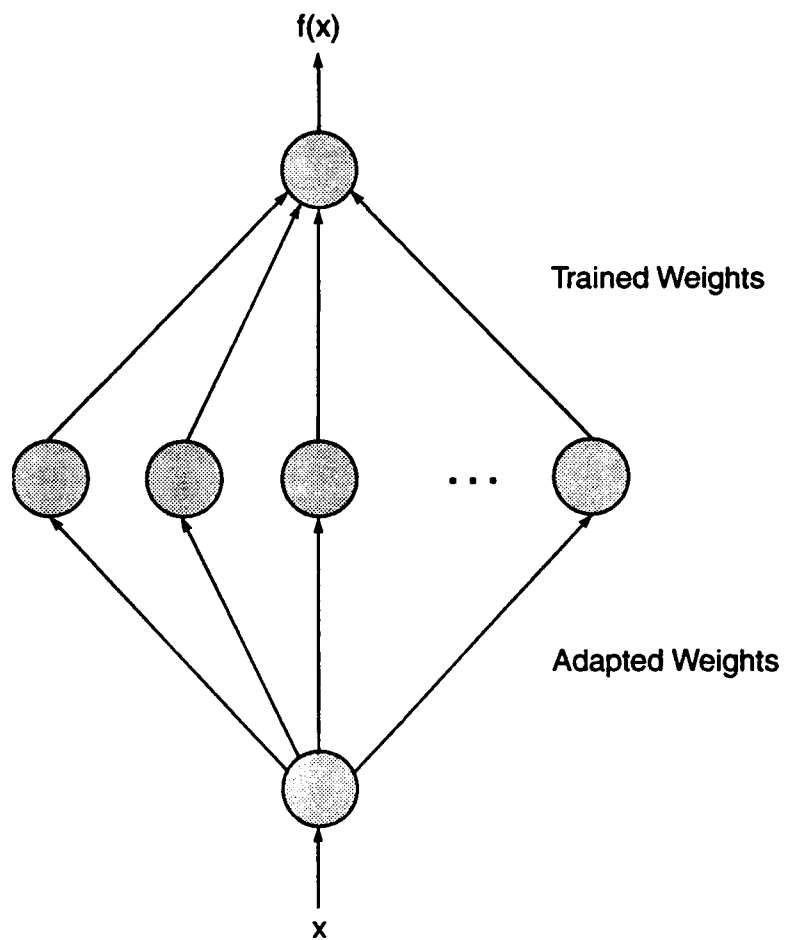


Figure 4. Single-Input, Single Output ANN Architecture

be represented by

$$f(t) = \sum_{m=-M}^M \sum_{n=0}^N c_{mn} \psi_{mn}(t) \quad (37)$$

where the $\psi_{mn}(t)$ represent the dilations and translations of the mother wavelet and the c_{mn} are the wavelet coefficients. The coefficients of the approximated signal outside its time-frequency space would be approximately zero since it would be outside the time and frequency support of the wavelet functions.

3.5 Summary

This chapter introduced the topic of WNNs by relating the areas of ANNs and wavelet analysis. Signal approximation can be performed with a SISO network with wavelet hidden-layer activation functions. The network weights from the hidden-layer to the output are learned and represent the projection coefficients of the signal being approximated. The next chapter shows the implementation of a particular mother wavelet composed of a summation of sigmoidal functions and the results of approximating signals with a hardware implementable WNN.

IV. RESULTS

4.1 Sigmoidal Wavelet

The first task at hand is to find a candidate mother wavelet function. The sigmoidal function is typically defined as the following:

$$s(t) = \frac{1}{1 + e^{-qt}}$$

A wavelet can be generated by combining sigmoids in the proper manner. Two sigmoidal functions are combined as follows

$$\phi(t) = s(t + d) - s(t - d) \quad (38)$$

$$= \frac{1}{1 + e^{-q(t+d)}} - \frac{1}{1 + e^{-q(t-d)}} \quad (39)$$

The resulting function, plotted in Figure 5 with $q = 2$ and $d = 1$, fails to meet the requirement of a mother wavelet having zero average energy.

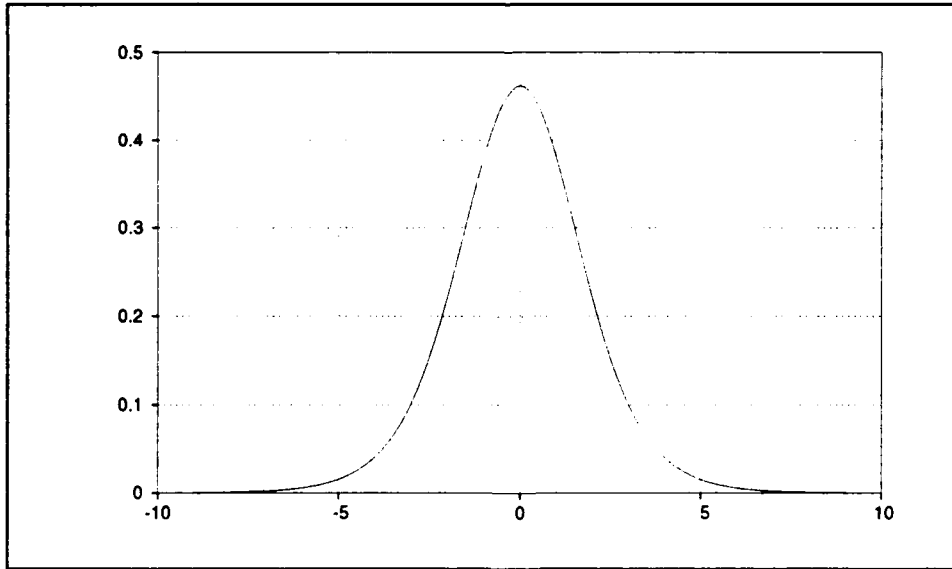


Figure 5. Combination of Two Sigmoidal Functions

By combining two ϕ -functions in the following manner, a candidate sigmoidal mother wavelet is obtained.

$$\psi(t) = \phi(t+p) - \phi(t-p) \quad (40)$$

and substituting with the sigmoid functions results in

$$\psi(t) = \frac{1}{1 + e^{-q(t-d-p)}} - \frac{1}{1 + e^{-q(t-d+p)}} - \frac{1}{1 + e^{-q(t+d-p)}} + \frac{1}{1 + e^{-q(t+d+p)}} \quad (41)$$

Letting $q = 2$, $d = 1$, and $p = 1$ gives three terms

$$\psi(t) = \frac{1}{1 + e^{-2(t-2)}} - \frac{2}{1 + e^{-2t}} + \frac{1}{1 + e^{-2(t+2)}} \quad (42)$$

The function ψ is shown in Figure 6. Obviously the function $\psi(t)$ has zero average energy and satisfies the conditions discussed in Equations 17 and 18 of Chapter 3 for the candidate mother wavelet.

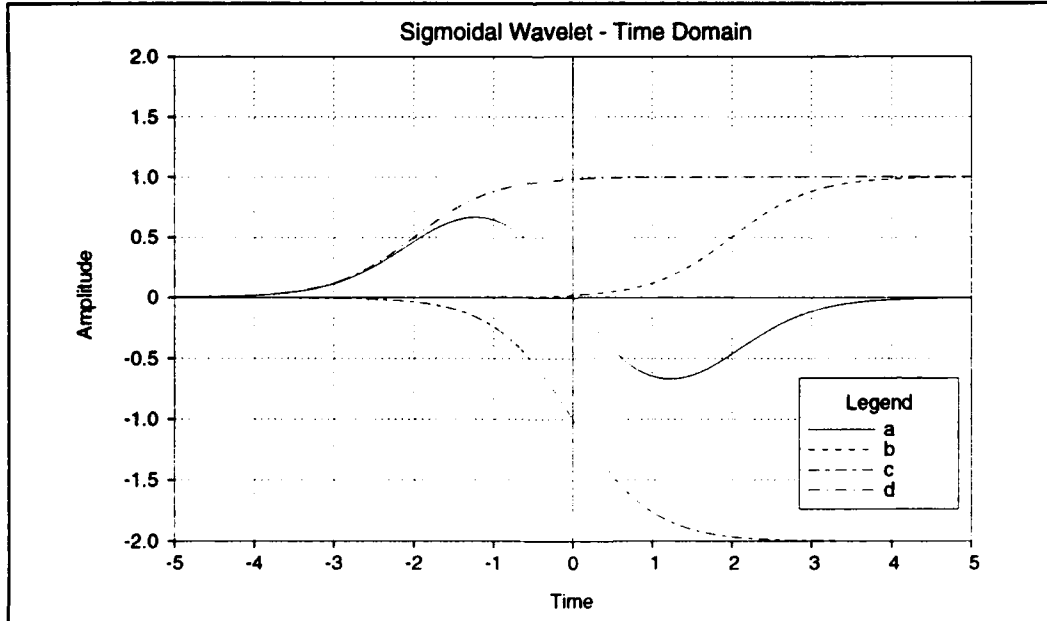


Figure 6. (a) Candidate Sigmoidal Mother Wavelet. (b) First term of Equation 42. (c) Second term of Equation 42. (d) Third term of Equation 42.

The Fourier Transforms of the functions ϕ and ψ can be calculated using complex analysis (see Appendix A for details) and results in

$$\Phi(\omega) = \int_{\mathbb{R}} \phi(t) e^{-j\omega t} dt = \frac{2\pi \sin(\omega d)}{q \sinh(\frac{\pi\omega}{q})} \quad (43)$$

$$\Psi(\omega) = e^{-jp\omega} \Phi(\omega) - e^{jp\omega} \Phi(\omega) \quad (44)$$

$$= \frac{-j4\pi}{q} \sin(\omega d) \sin(\omega p) \frac{1}{\sinh(\frac{\pi\omega}{q})} \quad (45)$$

The magnitude of the Fourier Transform of the candidate sigmoidal mother wavelet ψ is plotted in Figure 7. Note the shape of the Fourier Transform results in a bandpass filter as expected for discrete wavelets (6).

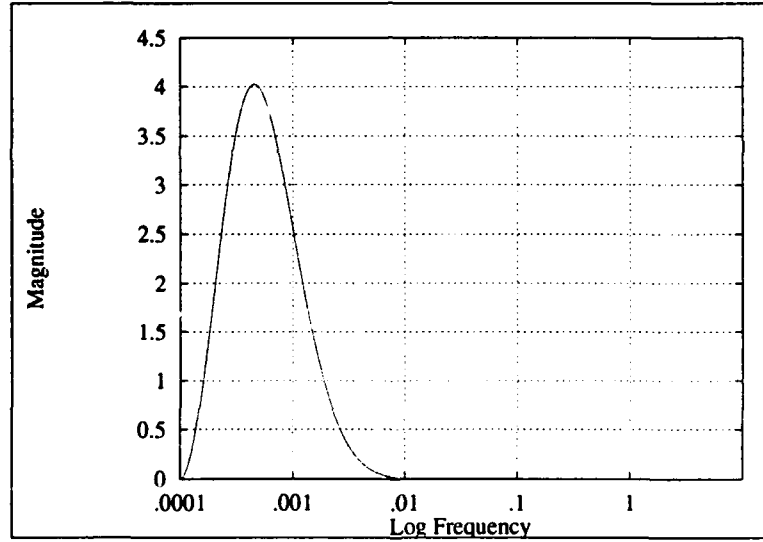


Figure 7. Fourier Transform of Candidate Sigmoidal Mother Wavelet, $\Psi(\omega)$

As pointed out in (11), a candidate mother wavelet does not necessarily result in a frame. The necessity to generate a frame occurs due to the requirement to “cover” the entire line in the time and frequency domains, therefore, sufficient translations must be used to meet the requirement for a frame. However, it is desirable to have as few translations as necessary to generate a frame because each translation, as will be seen,

results in an additional node in the WNN. The epsilon energy support of the sigmoidal mother wavelet, using Equation 23, is the smallest interval of $[t_0, t_1]$ which satisfies

$$\epsilon_t\text{-supp}(\psi) = \frac{1}{\|\psi\|^2} \int_{t_0}^{t_1} |\phi(t)|^2 dt < \epsilon_t \quad (46)$$

For $\epsilon_t = 0.1$, the epsilon energy support is $[-2.15, 2.15]$. In the frequency domain, $\Psi(\omega)$ is an even function and requires a somewhat different definition for the epsilon energy support. Recalling the epsilon energy support as defined in Equation 27 as

$$\epsilon_\omega\text{-supp}(\psi) = \frac{1}{2 \|\psi\|^2} \int_{\omega_0}^{\omega_1} |\Psi(\omega)|^2 d\omega < \epsilon_\omega \quad (47)$$

For an $\epsilon_\omega = 0.1$, the epsilon energy support of the sigmoidal mother wavelet is $[0.292, 1.592]$. With the epsilon support defined, it is possible to continue the task of defining a frame as outlined Chapter 3. The next step involves solving Equations 30 and 31 for the sigmoidal wavelet. Performing the calculations results in $m(\phi, a) = 7.88 > 0$ and $M(\phi, a) = 8.01 < 0$ (11). Therefore, a frame does exist, it is now a matter of selecting the value for b . The maximum value for this translation parameter was calculated in (11) to be $b = 0.57$.

4.2 Wavelet Neural Network

4.2.1 Network Architecture The required dilations and translations of the mother wavelet for representing a signal are determined by analyzing the area in the time-frequency space in which the signal lies. For a signal bandlimited to some range $[\omega_{min}, \omega_{max}]$ and existing over a finite interval $[t_{min}, t_{max}]$, the resulting area is as shown in Figure 8 (only positive frequencies are shown.) Choosing the appropriate dilations and translations to "cover" that area in the space results in the specifying the the range of dilations and translations. The bandwidth of the signal determines the number of dyadic dilations (constant-Q filters) required to represent the signal. Taking the ω_{max} and ω_{min} of the signal and dividing by the epsilon-support in the frequency domain of the mother wavelet

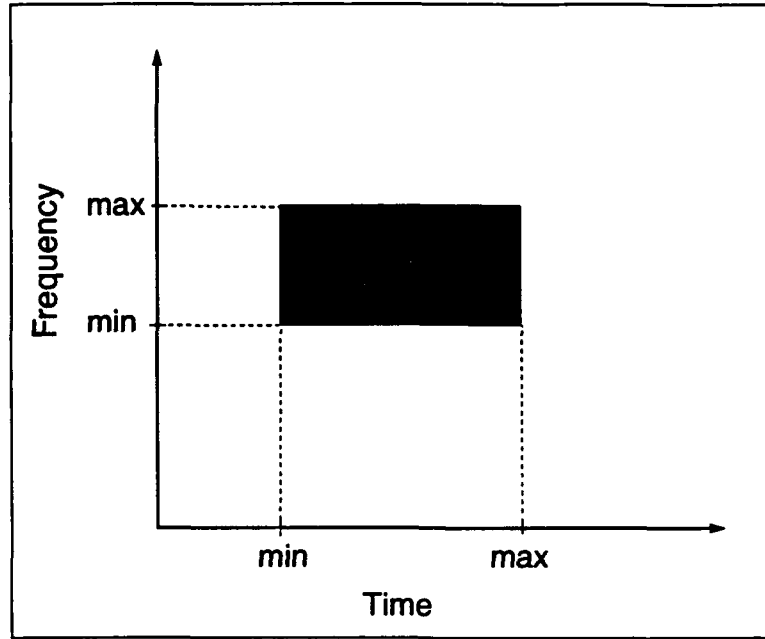


Figure 8. Time-Frequency Region for Approximation

provides the necessary information of what dilations are necessary along the frequency line.

$$2^{n_0} \omega_0(\psi) \leq \omega_{min} \quad (48)$$

$$2^{n_1} \omega_1(\psi) \geq \omega_{max} \quad (49)$$

Once the required dilations are determined, the range of frequencies covered by the dilations will be at least as big as the bandwidth of the signal, so

$$[2^{n_0} \omega_0(\psi), 2^{n_1} \omega_1(\psi)] \geq [\omega_{min}, \omega_{max}] \quad (50)$$

An example calculation follows: Given a signal limited to the frequency range of 5 to 15 Hz, determine the number of dilations of the sigmoidal mother wavelet required to fully

approximate the signal. For $\epsilon_\omega = 0.1$, then the epsilon-support is defined by the following

$$\omega_0(\psi) = 0.292 \quad (51)$$

$$\omega_1(\psi) = 1.592 \quad (52)$$

$$\omega_c(\psi) = 0.9420 \quad (53)$$

The bandwidth of the signal being approximated is 31.416 radians to 94.248 radians. Only one dilation is required for this bandwidth as shown in Table 1.

The method for determining the number of translations of the mother wavelet required is directly related to the dilations to be used and the interval of time desired for the approximation. The epsilon-support of the sigmoidal wavelet in the time domain is defined by

$$t_0(\psi) = -2.15 \quad (54)$$

$$t_1(\psi) = 2.15 \quad (55)$$

$$t_c(\psi) = 0.0 \quad (56)$$

For an approximation of 0.3 seconds duration of the signal referred to above where only the sixth dilation is required results in the center time and time span interval as shown in Table 2.

Table 1. Frequency Domain Parameters for Example

n	2^n	$2^n \omega_0(\psi)$	$2^n \omega_1(\psi)$	$2^n \omega_c(\psi)$
6	64	18.688	101.888	60.288

Table 2. Time Domain Parameters for Example

n	2^{-n}	$2^{-n} t_0(\psi)$	$2^{-n} t_1(\psi)$	$2^{-n} t_c(\psi)$
6	0.015625	-0.03359	0.03359	0

The number of hidden-layer nodes required for approximating a signal is determined by summing the number of translations for each dilation. Because dyadic dilations are used, it is also possible to calculate the total number of nodes using multiples of the number of translations of the smallest dilation. For example, if 10 translations of the 3rd dilation (smallest) are required, then 2 times the number of translations required for the 3rd dilation, or 20 translations, are required for the 4th dilation, 4 times the number translations for the 3rd dilation are required for the 5th dilation, and so on.

4.2.2 Software Implementation Implementing the Wave-Net in software involved changing the activation functions of a linear feed-forward ANN to represent the translated and dilated sigmoidal wavelet functions. This was accomplished by modifying code written in the C-language by D. Zahirniak and published in his master's thesis (17). Each node in the hidden layer was given an activation function to represent a particular dilation and translation of the mother wavelet.

4.2.3 Sampling of Training and Testing Data It is important to realize at this point that the output of a trained linear ANN can be calculated using matrix operations. In addition, it is possible to use matrix equations to explicitly solve for the coefficients if regularly spaced sampling of the function $f(t)$ are used (or assumed). Backpropagation training is a solution for finding the coefficients even if the sampling is irregularly spaced. Another advantage to the ANN representation of this approximation problem is that the sampling rate and intervals of the training data for learning the coefficients and the test sampling rates and intervals do not have to be the same. That is, the ANN could be trained using data sampled at one rate and accept an input for determining the approximation at a different sampling rate.

4.2.4 Approximation Results

4.2.4.1 Test Signals The signal approximation method using the sigmoid wavelet basis functions was first used to approximate a sum of two sinusoids to repeat

the results in (11) where

$$f(t) = \sin(2\pi 5t) + \sin(2\pi 10t) \quad (57)$$

was sampled over the period of 0 to 0.3 seconds. Using the range of frequencies between 5 and 10 Hertz requires the sixth dilation of the mother wavelet as discussed above. The number of translations required to represent this signal is the length of the signal divided by epsilon support of the 6th dilation of the mother wavelet. For this example, 40 dilations were required. Fifty randomly spaced samples of the function $f(t)$ were used to train the neural network.

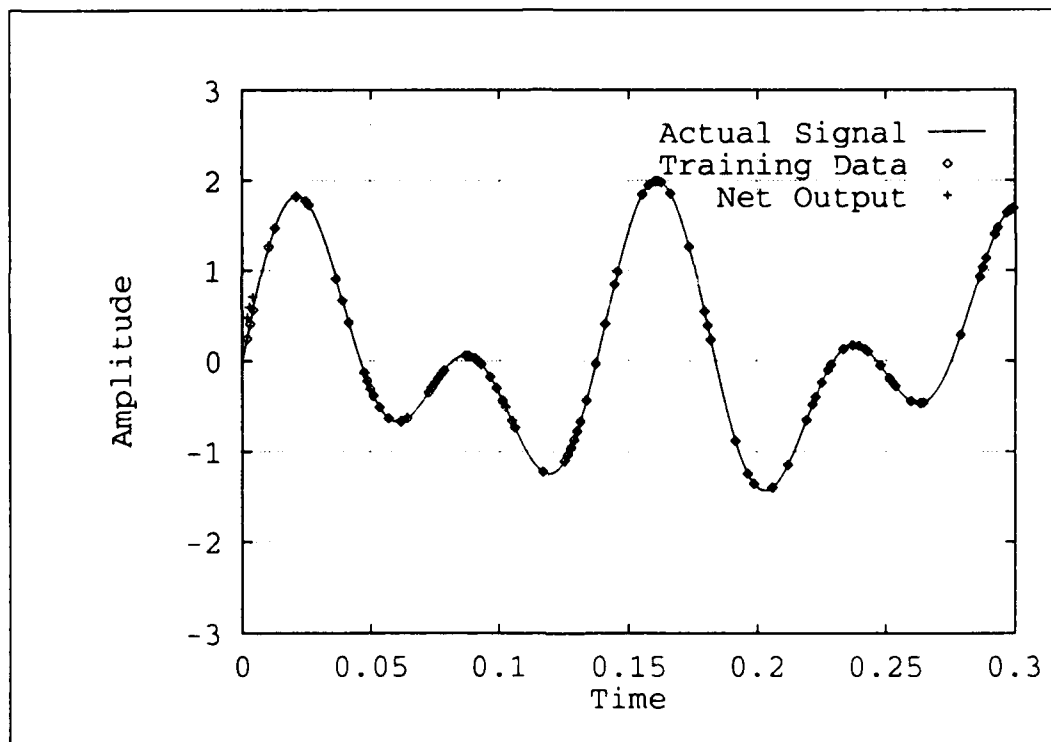


Figure 9. Approximation of Two Sinusoids

Further tests were conducted to verify the WNN's ability to learn sinusoids requiring different dilations of the mother wavelet as well as signals with a random spectrum limited to a chosen bandwidth. The random spectrum of the signal for this case was randomly

chosen between 5 and 200 Hz requiring the 6th, 7th, and 8th dilations of the mother wavelet. The approximation of this signal is shown in Figure 10.

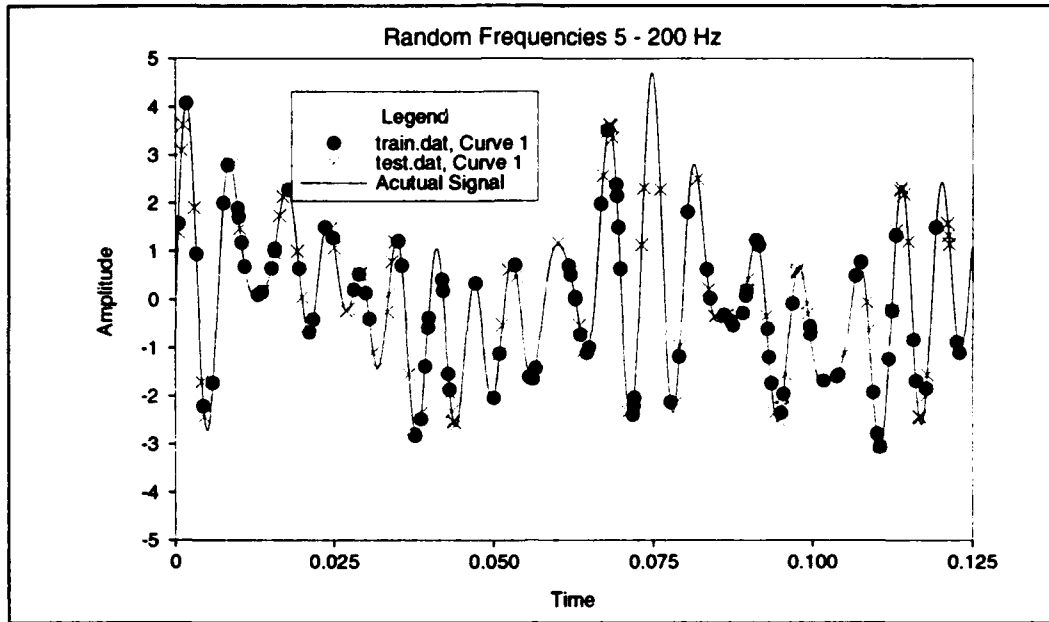


Figure 10. Approximation of Random Frequencies between 5 and 200 Hz

4.2.4.2 Real-World Signals The next step undertaken was to determine if the WNN could approximate real-world time-varying signals. An example of time sampled data obtained from the Wright Avionics Laboratory is shown in Figure 11. A Hilbert Transform routine in MATLAB was used to obtain the envelope of each signal resulting in signal pulses like that shown in Figure 12 where the samples of interest are highlighted. Further processing of the resulting envelope signal involved keeping only the first 64 samples of the envelope after a 3 dB threshold was exceeded. The Fourier Transform of the 64 samples was then used to determine the bandwidth of the envelope signal. The first approximation of the signal was chosen to include the first 12 of the 32 Fourier coefficients. The 12 Fourier coefficients corresponded to dilations of the mother wavelet from 0 to 6. The number of translations necessary to represent the signal resulted in 156 hidden nodes in the WNN such that 156 projection coefficients were learned.

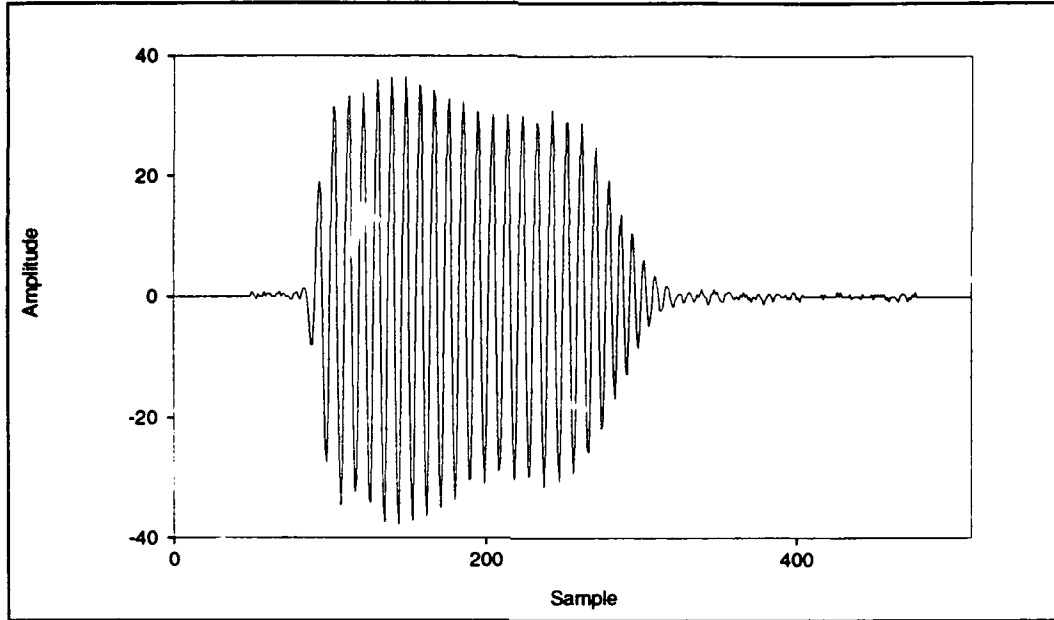


Figure 11. Sampled Signal of a Pulsed Carrier Frequency

The approximations appear to be low-pass filtered versions of the original envelope signal. As additional dilations of the mother wavelet are added to the WNN, higher frequencies components are approximated and the approximation more closely matches the original signal. As an example, including the 7th level of dilations in the WNN results in the approximation more closely matching the actual envelope than only including the 0 through 6 dilations as shown in Figure 13.

4.3 Hardware Implementation

The Intel ETANN chip uses a symmetric sigmoidal activation function that can be defined as the hyperbolic tangent function $\tanh(t)$. This function can also be used to generate a mother wavelet equivalent to the mother wavelet of Equation 39.

$$\tilde{s}(t) = \frac{2}{1 + e^{-qt}} - 1 \quad (58)$$

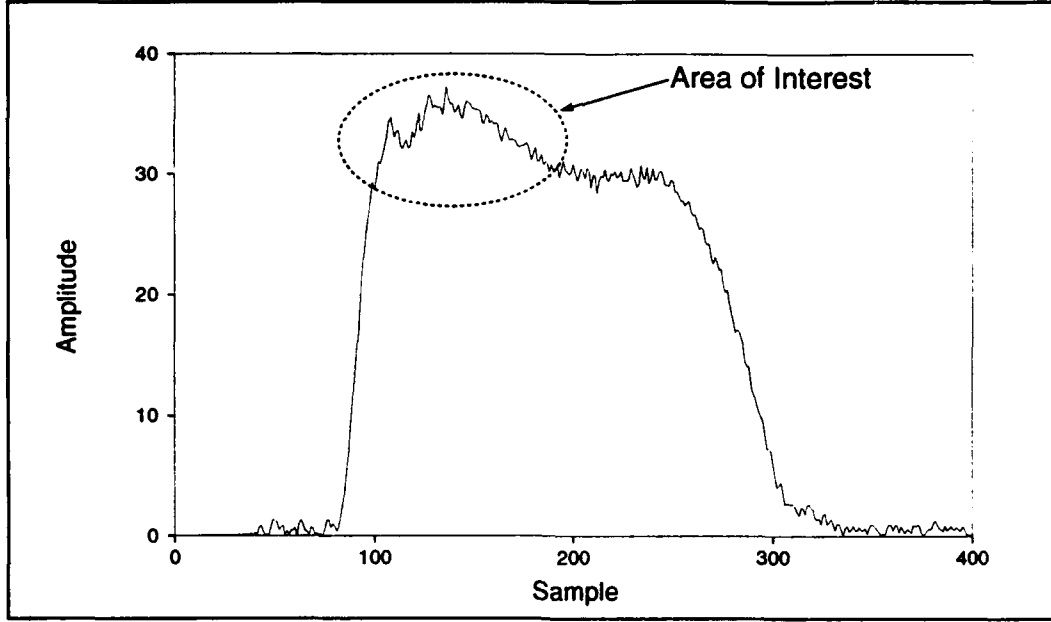


Figure 12. Envelope of Typical Signal After Hilbert Transform

with a wavelet function generated in a similar manner to Equation 41 resulting in

$$\tilde{\psi}(t) = \frac{2}{1 + e^{-qt}} * [\delta(t + d) - \delta(t - d)] * [\delta(t + p) - \delta(t - p)] \quad (59)$$

$$\tilde{\psi}(t) = 2\psi(t) \quad (60)$$

where $\psi(t)$ is defined in [41] and $\delta(t)$ is the dirac delta function. Using the Fourier Transform property that $cf(t) \xrightarrow{\mathcal{F}} cF(\omega)$ where c is an arbitrary constant, then

$$\tilde{\Psi}(\omega) = -j \frac{8\pi}{q} \sin \omega d \sin \omega p \frac{1}{\sinh \frac{\pi \omega}{q}} \quad (61)$$

The resulting function of Equation 59 has the same ϵ -energy support in the time and frequency domains as the mother wavelet developed in Equation 42 and all other calculations apply to the new wavelet function as well. Thus, the symmetric sigmoidal wavelet can be implemented in the ETANN hardware with no modifications to the previous development. The implementation just described still requires four hidden-layer nodes for each translation of each dilation required in the function approximation.

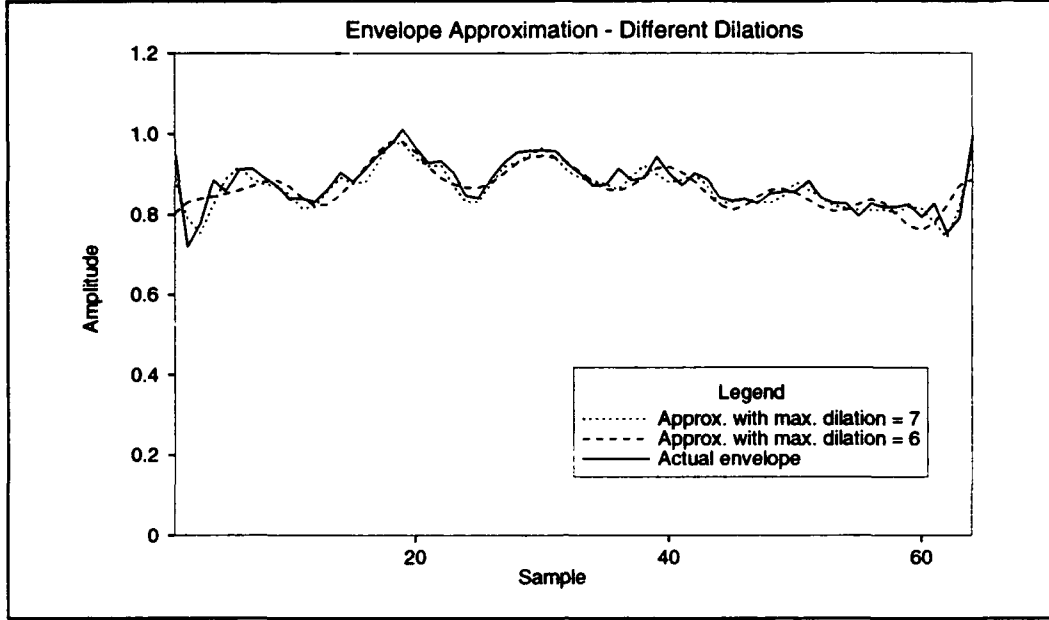


Figure 13. Approximation of Signal with Varying Numbers of Dilations

An alternative formulation for the mother wavelet , which uses only two hidden-layer nodes for each translation and dilation, can be represented by using

$$\hat{\psi}(t) = \left(\frac{2}{1 + e^{-q_1 t}} - 1 \right) - \left(\frac{2}{1 + e^{-q_2 t}} - 1 \right) \quad (62)$$

$$= \frac{2}{1 + e^{-q_1 t}} - \frac{2}{1 + e^{-q_2 t}} \quad (63)$$

which, letting $q_1 = 1$ and $q_2 = 2$, is plotted along with the original sigmoidal wavelet in Figure 14. Lowering the number of nodes is important because the ETANN chip has a maximum of 64 total nodes per chip in a single chip implementation and eight chips can be interconnected using the Intel development board for a total of 512 total nodes (3). The new mother wavelet function is slightly less concentrated in the time domain than the original mother wavelet, where the ϵ -energy support for $\epsilon_t = 0.1$ is $[-2.345, 2.345]$. The Fourier Transform for the wavelet of Equation 63 is

$$\hat{\Psi}(\omega) = -j2\pi \left(\frac{1}{q_1 \sinh \frac{\pi\omega}{q_1}} - \frac{1}{q_2 \sinh \frac{\pi\omega}{q_2}} \right) \quad (64)$$

which is plotted in Figure 15.

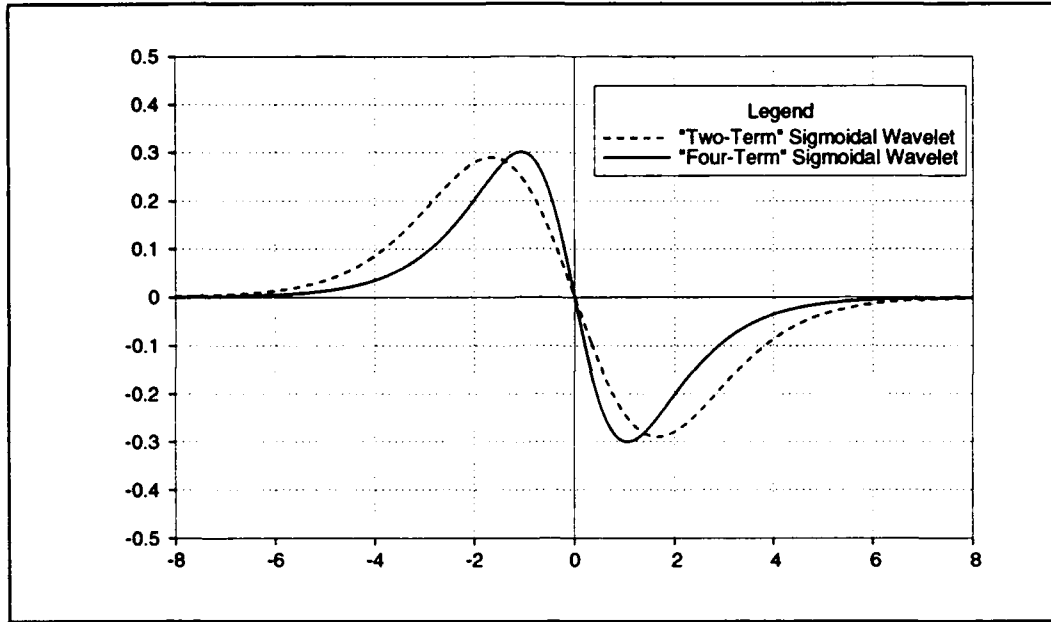


Figure 14. Sigmoidal Wavelet Functions Comparing the Symmetric Sigmoidal Wavelet of Equation 63 with the Sigmoidal Wavelet of Equation 42

Referring again to the SISO ANN described earlier in Figure 4, the hardware configuration would involve loading a set of weights to represent

$$f(t) = \sum_{m=-M}^M \sum_{n=0}^N c_{mn} \psi_{mn}(t) \quad (65)$$

$$\text{where } \psi_{mn}(t) = a^{\frac{n}{2}} \psi(a^n t - mb) \quad (66)$$

and the activation function of the output layer is linear. Note that t represents the single input. The equation representing the network configuration referenced to the output, say $g(t)$, of the hidden-layer nodes can be written as

$$f(t) = \sum_j w_{2,j} g_j(t) \quad (67)$$

where the subscript on w represents the j th node in the second layer of the network. The output of the hidden-layer can be represented in terms of the network input, the first layer

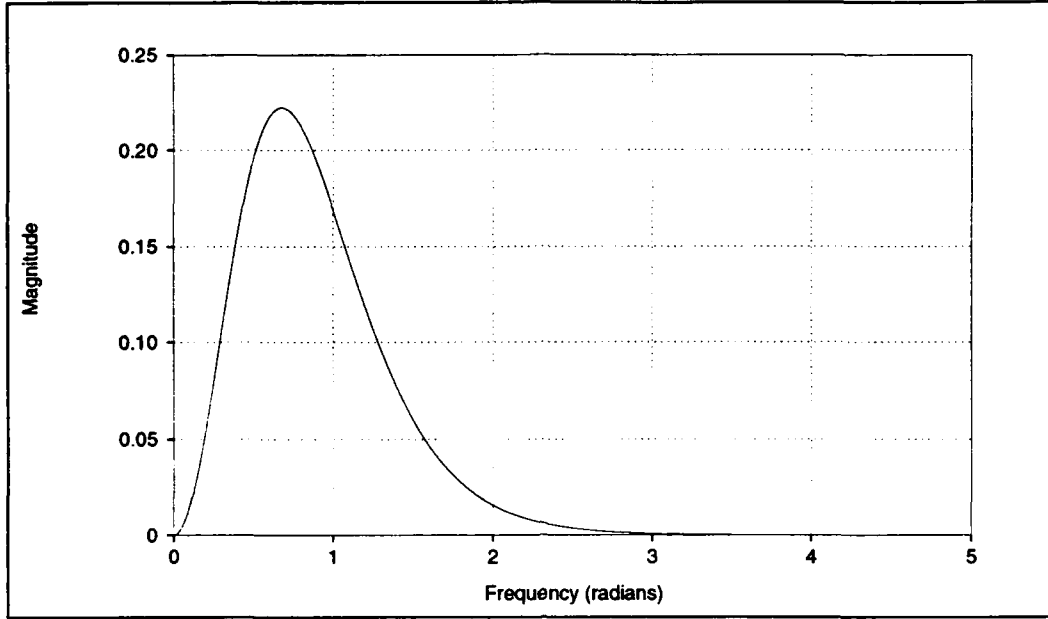


Figure 15. Magnitude of FT of Symmetric Sigmoidal Wavelet

weights, and the hidden-layer activation function, that is,

$$g_j(t) = h(w_{1,j}t + \beta_j) \quad (68)$$

where $h(\cdot)$ is the activation function of the hidden layer and β_j is the bias of the j th hidden node. Equating terms from Equations 66 and 68 yields

$$mb = \beta_j \quad (69)$$

$$a^n t = w_{1,j}t \quad (70)$$

$$\psi(\cdot) = h(\cdot) \quad (71)$$

Therefore, $w_{1,j}$ represent the dilations and β_j represent the translations of the mother wavelet. Because ψ represents the summation of four sigmoidal functions, it is necessary to add an additional layer of nodes to the SISO network of Figure 4. Letting the weights from second hidden layer to the output layer be $a^{\frac{n}{2}}$ and setting the weights between the

hidden layers to 1 and -1 as shown in Figure 16 results in a WNN architecture that can be implemented in hardware like the ETANN.

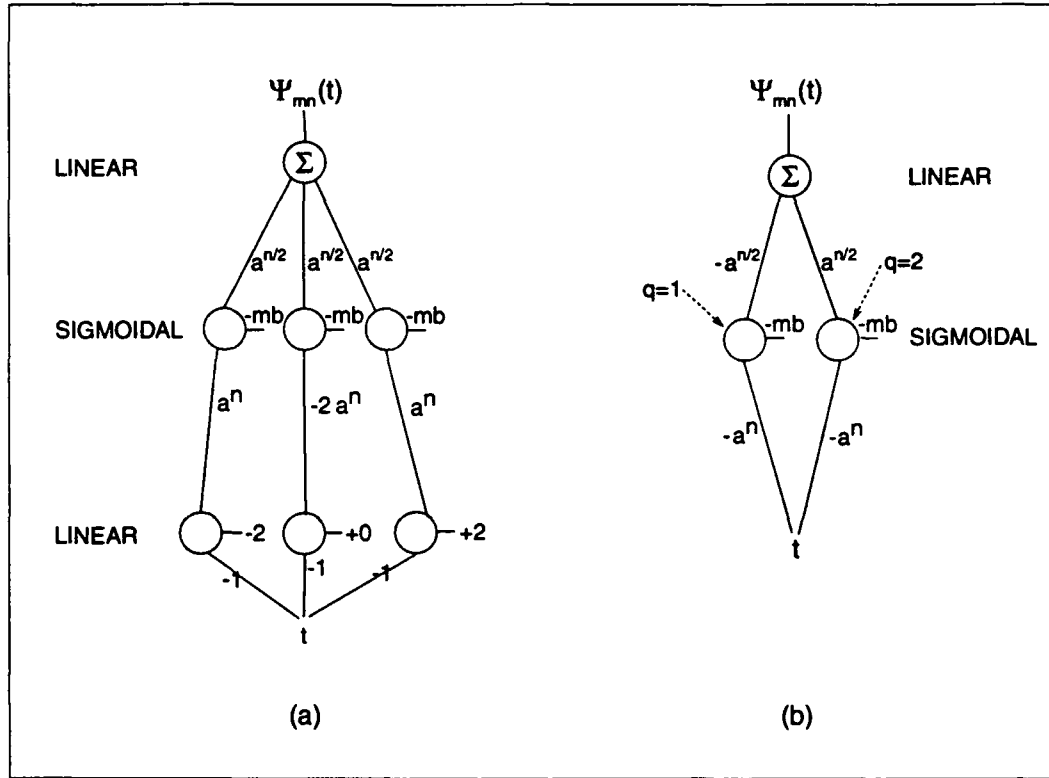


Figure 16. Hardware Implementation for Sigmoidal Wavelets Illustrating Reduced Number of Hidden-Layer Nodes for the Two-Term Sigmoidal Wavelet Representation: (a) Wavelet from Summation of Three Sigmoidal Functions, (b) Wavelet from Two Symmetric Sigmoidal Functions

V. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The results of this research demonstrated:

- The sigmoidal wavelet is useful for representing real-world bandlimited signals.
- The Wavelet Neural Network (WNN) can learn to approximate bandlimited signals.

The research also verified the following conclusions:

- The wavelet representation of an ANN for approximating a function allow the number of required nodes to be calculated explicitly.
- The weights from the input layer to the hidden layer can be fixed once the time-frequency space in which the function resides is determined. This property greatly reduces the number of weights that must be learned.
- Non-uniformly sampled data can be used to train the WNN to achieve good approximations.

5.2 Recommendations

Several areas in which this research topic could continue are

- Using the learned projection coefficients to experiment with a large class of signals.
- Investigate neural networks (or some other real-time or near real-time method) for determining the projection coefficients directly rather than having to learn them. This capability would allow the coefficients to be loaded into a feedforward ANN for classification.
- A problem that requires the approximation and/or interpolation of another bandlimited signal. This could also include three-dimensional functions.

Appendix A. *Fourier Transforms of Sigmoidal Wavelets*

A sum of four sigmoids generates a wavelet function as discussed in Section 4.1. This appendix provides the derivation of the Fourier Transform of the resulting wavelet which is required to calculate the epsilon-energy support in the frequency domain. The sigmoid function is defined as

$$s(t) = \frac{1}{1 + e^{-qt}} \quad (72)$$

Subtracting two shifted sigmoid functions gives

$$\phi(t) = s(t + d) - s(t - d) \quad (73)$$

$$= \frac{1}{1 + e^{-q(t+d)}} - \frac{1}{1 + e^{-q(t-d)}} \quad (74)$$

which can be written as

$$\phi(t) = \frac{1}{1 + e^{-qt}} * [\delta(t + d) - \delta(t - d)] \quad (75)$$

where “*” indicates the convolution operation. Writing the time-shifted sigmoids as a single sigmoid function convolved with delta functions simplifies the derivation of the Fourier Transform that follows. A wavelet function is generated from Equation 75 by letting

$$\begin{aligned} \psi(t) &= \phi(t + p) - \phi(t - p) \\ &= \phi(t) * [\delta(t + p) - \delta(t - p)] \\ &= \frac{1}{1 + e^{-qt}} * [\delta(t + d) - \delta(t - d)] * [\delta(t + p) - \delta(t - p)] \end{aligned}$$

Letting $q = 2$ and $d = p = 1$ results in

$$\psi(t) = \frac{1}{1 + e^{-2(t-2)}} - \frac{2}{1 + e^{-2t}} + \frac{1}{1 + e^{-2(t+2)}} \quad (76)$$

Taking the Fourier Transform of $\psi(t)$

$$\Psi(\omega) = \int_{-\infty}^{+\infty} \psi(t) e^{-j\omega t} dt \quad (77)$$

and using the Fourier Transform property that $h(t) * g(t) \xrightarrow{\mathcal{F}} H(\omega)G(\omega)$ results in

$$\Psi(\omega) = -4 \sin(\omega d) \sin(\omega p) \int_{-\infty}^{+\infty} \frac{1}{1 + e^{-qt}} e^{-j\omega t} dt \quad (78)$$

Notice that taking only the integral part of the above equation results in the Fourier Transform of the original sigmoid function given in (72).

$$S(\omega) = \int_{-\infty}^{+\infty} \frac{1}{1 + e^{-qt}} e^{-j\omega t} dt \quad (79)$$

The function $S(\omega)$ cannot easily be evaluated using the usual methods of real integral calculus, thus, complex analysis is needed. To evaluate the integral, first it is written as

$$\int_{-R}^{+R} \frac{1}{1 + e^{-qt}} e^{-j\omega t} dt = \oint_{C_R} \frac{1}{1 + e^{-qz}} e^{-j\omega z} dz - \int_{S_R} \frac{1}{1 + e^{-qz}} e^{-j\omega z} dz \quad (80)$$

where $C_R = [-R, R] \cup S_R$ is a closed contour transversed in a counterclockwise direction and S_R is the semicircle in the upper-half plane with radius R centered at the origin as shown in Figure 17.

If the integrand of the first integral of Equation 80 has singularities at points inside C_R , but is otherwise analytic, then the integral can be evaluated using the Residue Integration Method (8). Substituting $z = -R(\cos \theta + j \sin \theta) = -Re^{j\theta}$ for $z \in S_R$ into the second integral yields

$$-\int_{S_R} \frac{e^{-j\omega z}}{1 + e^{-qz}} dz = \int_0^\pi \frac{e^{j\omega R \cos \theta} e^{-\omega R \sin \theta}}{1 + e^{qR \cos \theta} e^{jqR \sin \theta}} Re^{j\theta} j d\theta \quad (81)$$

Observe that

$$\left| \int_0^\pi \frac{e^{j\omega R \cos \theta} e^{-\omega R \sin \theta}}{1 + e^{qR \cos \theta} e^{jqR \sin \theta}} Re^{j\theta} j d\theta \right| \leq \int_0^\pi \frac{e^{-\omega R \sin \theta} R}{|1 + e^{qR \cos \theta} e^{jqR \sin \theta}|} d\theta \quad (82)$$

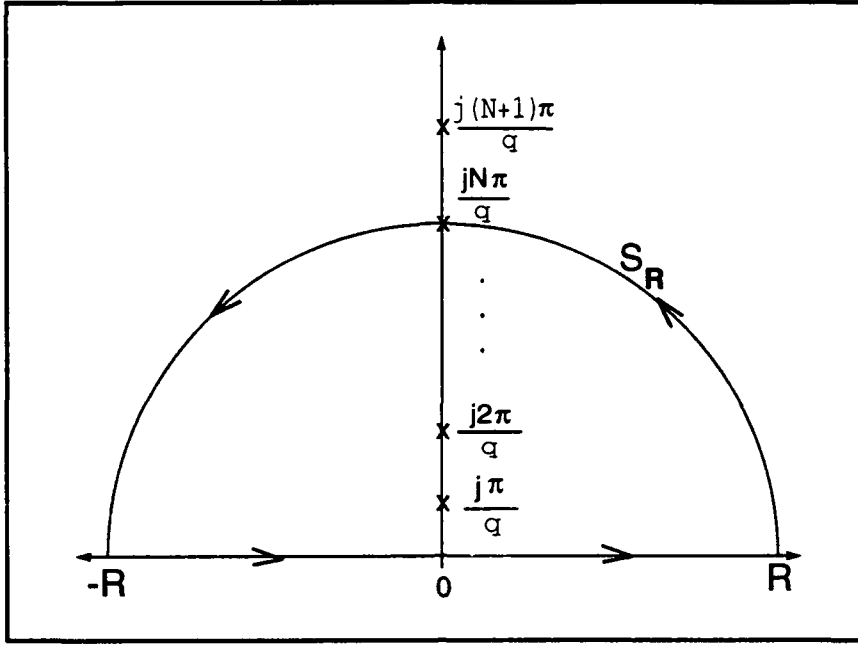


Figure 17. Integration Contour

On the interval $0 < \theta < \pi/2$, $0 < \cos \theta < 1$ so $e^{qR \cos \theta} > 1$ thus

$$|1 + e^{qR \cos \theta} e^{jqR \sin \theta}| \geq e^{qR \cos \theta} - 1 > 0 \quad (83)$$

On the interval $\pi/2 < \theta < \pi$, $-1 < \cos \theta < 1$ so $e^{qR \cos \theta} < 1$ thus

$$|1 + e^{qR \cos \theta} e^{jqR \sin \theta}| \geq 1 - e^{qR \cos \theta} > 0 \quad (84)$$

Therefore,

$$\int_0^\pi \frac{e^{-\omega R \sin \theta} R}{|1 + e^{qR \cos \theta} e^{jqR \sin \theta}|} d\theta \leq \int_0^{\pi/2} \frac{e^{-\omega R \sin \theta} R}{e^{qR \cos \theta} - 1} d\theta + \int_{\pi/2}^\pi \frac{e^{-\omega R \sin \theta} R}{1 - e^{qR \cos \theta}} d\theta \quad (85)$$

Taking the limit as $R \rightarrow \infty$ yields that both integrals converge to zero. The resulting integral to be evaluated is

$$\oint_{C_R} \frac{e^{-j\omega z}}{1 + e^{-qz}} dz \quad (86)$$

Replacing z by $-\zeta$ gives

$$\oint_{C_R} \frac{e^{-j\omega z}}{1 + e^{-qz}} dz = \oint_{-C_R} \frac{e^{j\omega\zeta}}{1 + e^{q\zeta}} d\zeta = \oint_{C_R} \frac{e^{j\omega\zeta}}{1 + e^{q\zeta}} d\zeta \quad (87)$$

where $-C_R$ is the contour C_R transversed in the clockwise direction. Replacing ζ by z results in

$$f(z) = \frac{e^{j\omega z}}{1 + e^{qz}} \quad (88)$$

Because multiple singularities of $f(z)$ are contained within the contour of integration, the Residue Theorem states (8),

$$\oint_C f(z) dz = j2\pi \sum_{k=1}^n \text{Res}_{z=z_k} f(z) \quad (89)$$

To use the Residue Theorem, it is first necessary to find the singularities, i.e., where $f(z)$ is not analytic which occur when

$$1 + e^{qz} = 0 \quad (90)$$

or, equivalently, when $e^{qz} = -1$. Solving for z , we obtain

$$\begin{aligned} qz &= j\pi(2n+1) \text{ for } n \in \mathcal{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\} \\ z &= \frac{j\pi}{q}(2n+1) \text{ for } n \in \mathcal{Z} \end{aligned} \quad (91)$$

Since the integration contour as shown in Figure 17 is over the upper-half plane, then only positive integers of n are contained within the closed contour resulting in

$$z_n = \frac{j\pi}{q}(2n+1), n \in \mathcal{Z}^+ = \{0, 1, 2, \dots\} \quad (92)$$

Residues from singularities inside the integration contour are included in the summation when calculating the residues.

$$\lim_{R \rightarrow \infty} \int_{-R}^{+R} f(z) dz = \lim_{N \rightarrow \infty} \int_{-2\pi N/q}^{+2\pi N/q} f(z) dz = \lim_{N \rightarrow \infty} j2\pi \sum_{n=0}^{N-1} \text{Res}_{z=z_n} f(z) \quad (93)$$

The particular $f(z)$ for this case is of the form

$$f(z) = \frac{g(z)}{h(z)} \quad (94)$$

where $g(z) \neq 0$ and $h(z)$ has a simple zero at $z = z_0$ ($h(z_0) = 0$ and $h'(z_0) \neq 0$). Therefore, z_0 is a simple pole of $f(z)$ and

$$\text{Res}_{z=z_0} f(z) = \lim_{z \rightarrow z_0} \frac{(z - z_0)g(z)}{h(z)} = \frac{g(z_0)}{h'(z_0)} \quad (95)$$

Substituting for $g(z_0)$ and $h(z_0)$ results in

$$\lim_{N \rightarrow \infty} j2\pi \sum_{n=0}^{N-1} \frac{e^{j\omega z_n}}{q e^{q z_n}} = j2\pi \sum_{n=0}^{\infty} \frac{e^{j\omega \frac{j\pi(2n+1)}{q}}}{q e^{j\pi(2n+1)}} \quad (96)$$

Rearranging and recalling that $e^{j\pi(2n+1)} = -1$ results in

$$\int_C f(z) dz = -\frac{j2\pi}{q} \sum_{n=0}^{\infty} e^{-(2n+1) \frac{\pi\omega}{q}} \quad (97)$$

From the CRC Standard Mathematical Tables (2),

$$\sinh a = \left(2 \sum_{n=0}^{\infty} e^{(-2n+1)a} \right)^{-1} \text{ for } \Re(a) > 0 \quad (98)$$

Therefore, the Fourier Transform of the sigmoidal function is

$$\int_C f(z) dz = -\frac{j\pi}{q} \frac{1}{\sinh \frac{\pi\omega}{q}} \text{ for } \frac{\pi\omega}{q} > 0 \quad (99)$$

and the Fourier Transform of the sigmoidal wavelet is

$$\begin{aligned}\Psi(\omega) &= (-4 \sin(\omega d) \sin(\omega p)) \left(-\frac{j\pi}{q} \frac{1}{\sinh \frac{\pi\omega}{q}} \right), \frac{\pi\omega}{q} > 0 \\ &= \frac{-j4\pi}{q} \sin \omega d \sin \omega p \frac{1}{\sinh \frac{\pi\omega}{q}}\end{aligned}\quad (100)$$

The Fourier Transform of a wavelet generated by symmetric sigmoids can be derived from Equation 100. Recall the symmetric sigmoid is defined as

$$\tilde{s}(t) = \frac{2}{1 + e^{-qt}} - 1 \quad (101)$$

with a wavelet function generated in a similar manner to Equation 76 resulting in

$$\tilde{\psi}(\omega) = \frac{2}{1 + e^{-qt}} * [\delta(t + d) - \delta(t - d)] * [\delta(t + p) - \delta(t - p)] \quad (102)$$

Using the Fourier Transform property that $cf(t) \xrightarrow{\mathcal{F}} cF(\omega)$ where c is an arbitrary constant, then

$$\tilde{\Psi}(\omega) = -j \frac{8\pi}{q} \sin \omega d \sin \omega p \frac{1}{\sinh \frac{\pi\omega}{q}} \quad (103)$$

A wavelet generated from only two symmetric sigmoids by varying the factor q is given by

$$\hat{\psi}(t) = \frac{2}{1 + e^{-q_1 t}} - \frac{2}{1 + e^{-q_2 t}} \quad (104)$$

For this case, the Fourier Transform is

$$\hat{\Psi}(\omega) = -2j\pi \left(\frac{1}{q_1 \sinh \frac{\pi\omega}{q_1}} - \frac{1}{q_2 \sinh \frac{\pi\omega}{q_2}} \right) \quad (105)$$

Bibliography

1. Bakshi, Bhavik R. and George Stephanopoulos. "Wavelets as Basis Functions for Localized Learning in a Multi-Resolution Hierachy," *International Joint Conference on Neural Networks, II*:140-145 (June 1992).
2. Beyer, William H., editor. *CRC Standard Mathematical Tables* (27th Edition). CRC Press, 1984.
3. Calvin, James B. *ETANN Hardware Implementation for Radar Emitter Identification*. MS thesis, Air Force Institute of Technology, December 1992.
4. Cameron, David M. *Radar System Classification Using Neural Networks*. MS thesis, Air Force Institute of Technology, December 1991.
5. Cybenko, G. *Approximations by Superpositions of a Sigmoidal Function*. Technical Report 856. Technical Report, Center for Supercomputing Research and Development, University of Illinois, Champaign-Urbana, IL, February 1989.
6. Daubechies, Ingrid. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.
7. Gaskill, Jack. *Linear Systems, Fourier Transforms, and Optics*. John Wiley & Sons, 1978.
8. Kreyszig, Erwin. *Advanced Engineering Mathematics, 6th Edition*. John Wiley & Sons, Inc., 1988.
9. Mallat, Stephane G. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674-693 (1989).
10. Oppenheim, Allan and R. Schafer. *Discrete-Time Signal Processing*. John Wiley & Sons, Inc., 1991.
11. Pati, Y. C. and P. S. Krishnaprasad. *Analysis and Synthesis of Feedforward Neural Networks Using Discrete Affine Wavelet Transformations*. Technical Report, Technical research report of the Electrical Engineering Department and Systems Research Center. University of Maryland, College Park. MD 20742, TR 90-44.
12. Rioul, Olivier and Martin Vetterli. "Wavelets and Signal Processing," *IEEE Signal Proc Magazine* (October 1991).
13. Rogers, Steven K. and Matthew Kabrisky. *An Introduction to Biological and Artificial Neural Networks for Pattern Recognition*. SPIE Optical Engineering Press, 1991.
14. Ruck, Dennis. et al. "The multilayer perceptron as an approximation to a Bayes Optimal discriminant function," *IEEE Transactions on Neural Networks*, 1:296-299 (December 1990).
15. Schalkoff, Robert J. *Pattern Recognition: Statistical, Structural and Neural Approaches*. John Wiley & Sons, Inc., 1992.

16. Vetterli, Martin and Cormac Herley. "Wavelets and Filter Banks: Theory and Design," *IEEE Transactions on Information Theory*, 38(2) (March 1992).
17. Zahirniak, Daniel R. *Characterization of Radar Signals Using Neural Networks*. MS thesis, Air Force Institute of Technology, December 1990.

Vita

Charles (Chuck) Westphal graduated from high school in 1979 from Johannesburg-Lewiston H.S., Johannesburg, Michigan. He then attended Michigan Technological University where he received a B.S. degree in chemical engineering in 1983. After completing Officer Training School, Chuck was commissioned as a second lieutenant in the U.S. Air Force in 1984. He received a B.S. in electrical engineering from the University of New Mexico in 1986. From 1987 to 1991, he served as an electronic warfare engineer and flight test program manager for F-16 electronic combat systems at the USAF Tactical Air Warfare Center, Eglin AFB, FL. In June of 1991, he began his pursuit of a master's degree in electrical engineering at the Air Force Institute of Technology. Chuck is currently employed with MacAulay Brown Inc of Dayton Ohio after separating from the Air Force in 1992.

Permanent address: 1480 Snider Road,
New Carlisle, OH 45344

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE December 1992	3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE SIGNAL APPROXIMATION WITH A WAVELET NEURAL NETWORK			5. FUNDING NUMBERS	
6. AUTHOR(S) Charles M. Westphal				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology, WPAFB OH 45433-6583			8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/GE/ENG/92D-38	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Captain Daniel Zahirniak WDRC/AAWP-1 WPAFB, OH 45433-6543			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Distribution Unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This study investigated the use of Wavelet Neural Networks (WNN) for signal approximation. The particular wavelet function used in this analysis consisted of a summation of sigmoidal functions (a <i>sigmoidal wavelet</i>). The sigmoidal wavelet has the advantage of being easily implemented in hardware via specialized electronic devices like the Intel Electronically Trainable Analog Neural Network (ETANN) chip. The WNN representation allows the determination of the number of hidden-layer nodes required to achieve a desired level of approximation accuracy. Results show that a bandlimited signal can be accurately approximated with a WNN trained with irregularly sampled data.				
14. SUBJECT TERMS Signal Approximation, Wavelet Neural Network			15. NUMBER OF PAGES 57	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	